RapTOR: Automated sequencing library preparation and suppression for rapid pathogen characterization

Todd W. Lane, Principal Investigator

SFAF Meeting, June 7, 2012

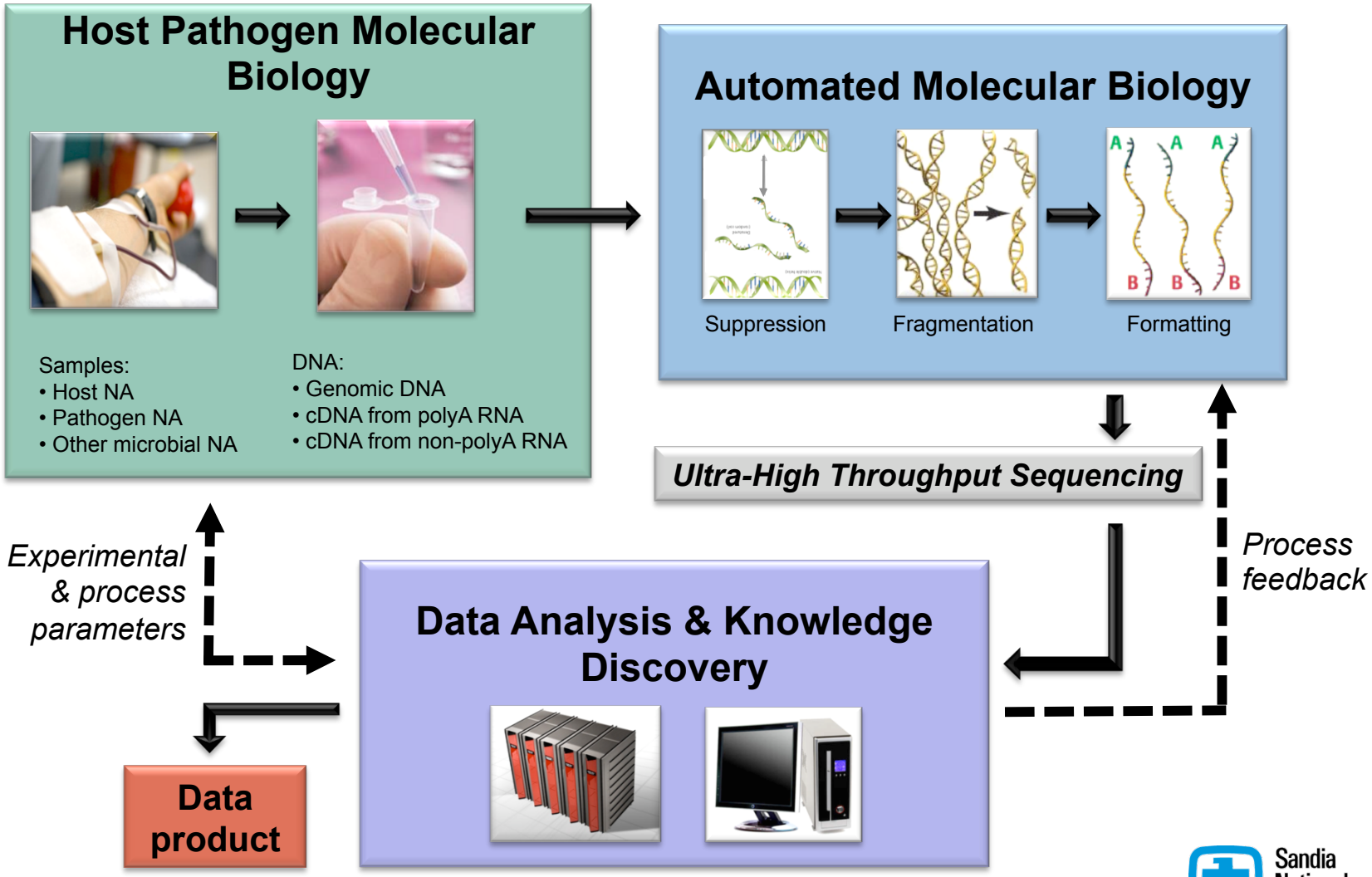
Santa Fe, NM

LDRD
LABORATORY DIRECTED RESEARCH & DEVELOPMENT

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

RapTOR system concept

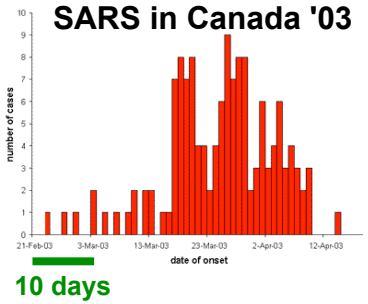


Sandia National Laboratories

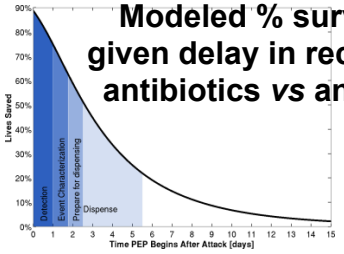
RapTOR is designed to address large outbreak scenarios

Outbreak dynamics are measured in ***days to weeks***.

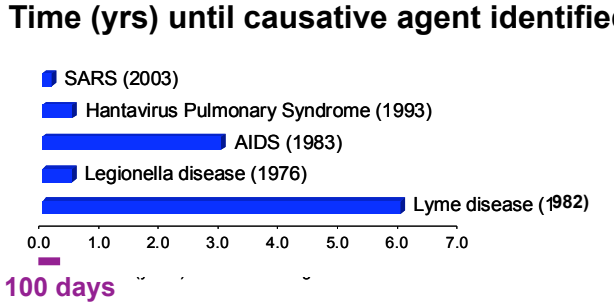
Culture-based identification of agent can take ***months to years***.



10 days



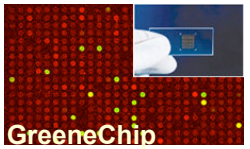
10 days

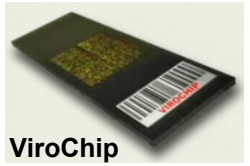


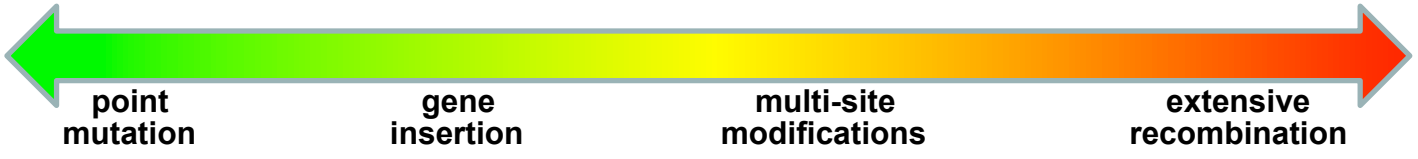
100 days

Probe-based methods are fast but often confounded.

- Don't recognize deeply diverged targets.
- Blind to unanticipated features (*eg*, IL-4).
- Unusual profiles are difficult to interpret.
  - Mix-and-match features? Artifact?







Second Generation Sequencing methods are transformational for pathogen characterization

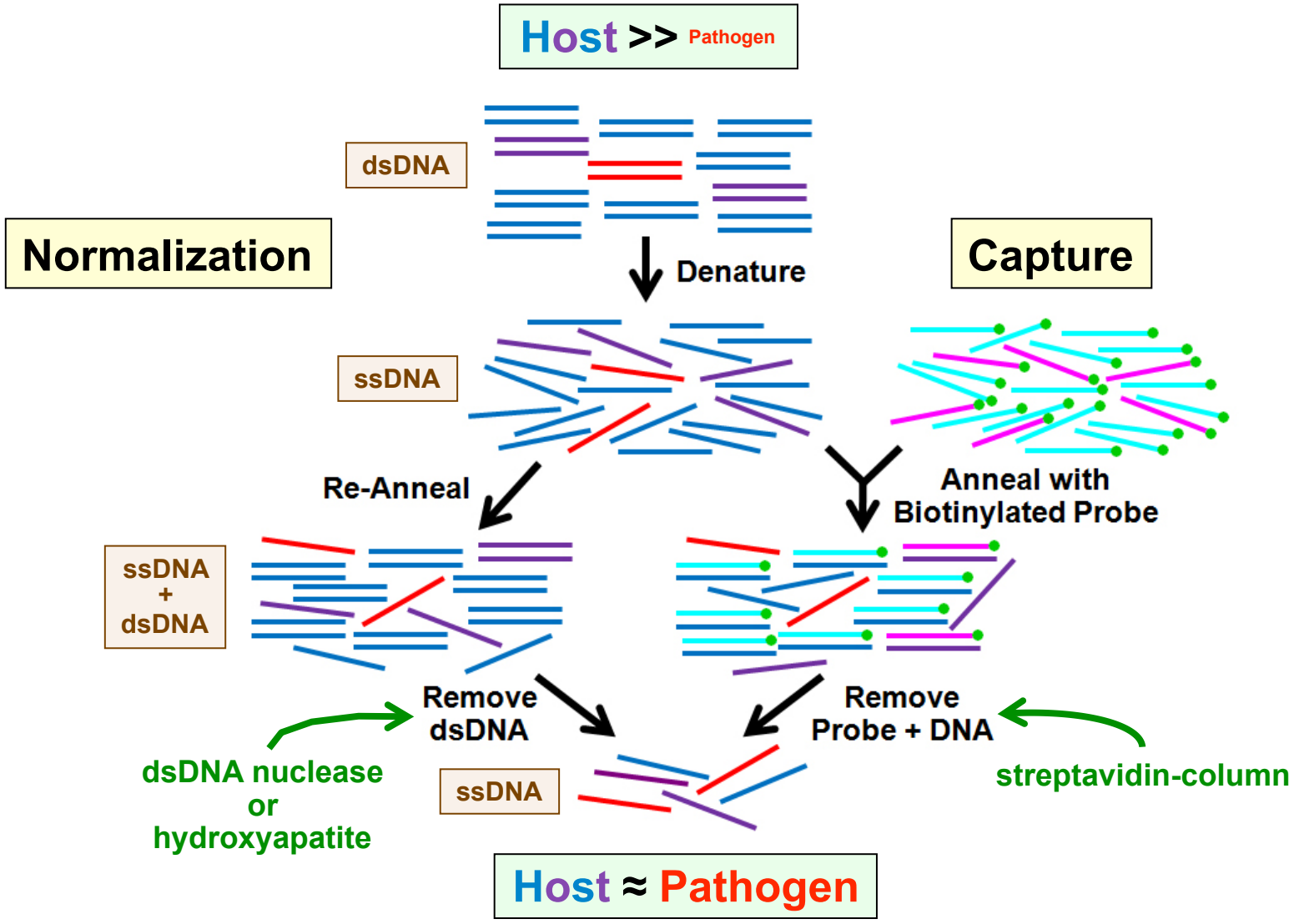
Brute-force SGS of clinical samples has enabled the discovery of novel pathogens

http://www.illumina.com/systems/hiseq_2000.ilmn

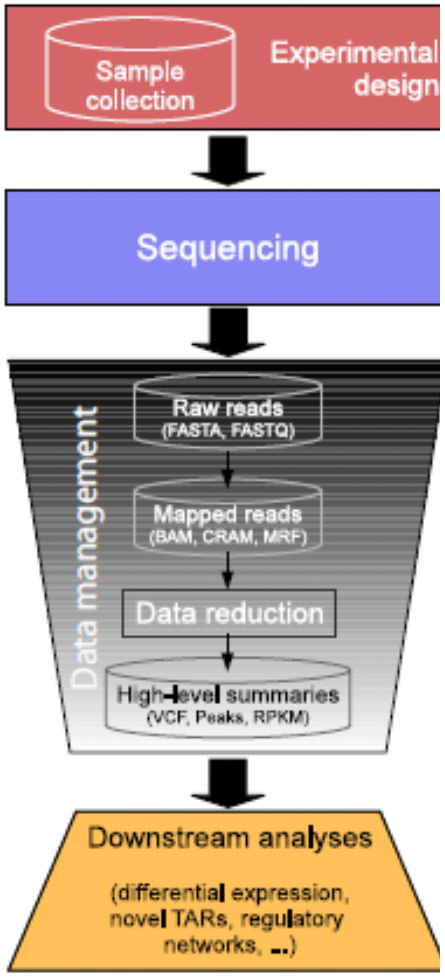
Disease	Sample	Novel Agent Detected	Total Reads	Hits on Agent	Reference	Suppression for 1X coverage
Merkel cell carcinoma	tumors	"Merkel cell polyomavirus"	395,734	2 (**0.00005%**)	Science 319:1096 '08	14-18X
organ transplant related fatality	serum & organs	"Dandenong" arenavirus	103,632	14 (**0.014%**)	N Engl J Med 358:991 '08	5X
pediatric gastroenteritis	feces	"human klassevirus "	937,935	849 (**0.09%**)	Virol J 6:82 '09	0 (75%)
pediatric influenza-like illness	nasopharyngeal swabs	"human enterovirus type 109"	20,825,810	119 (**0.0006%**)	J Virol 84:9047 '10	0 (40%)

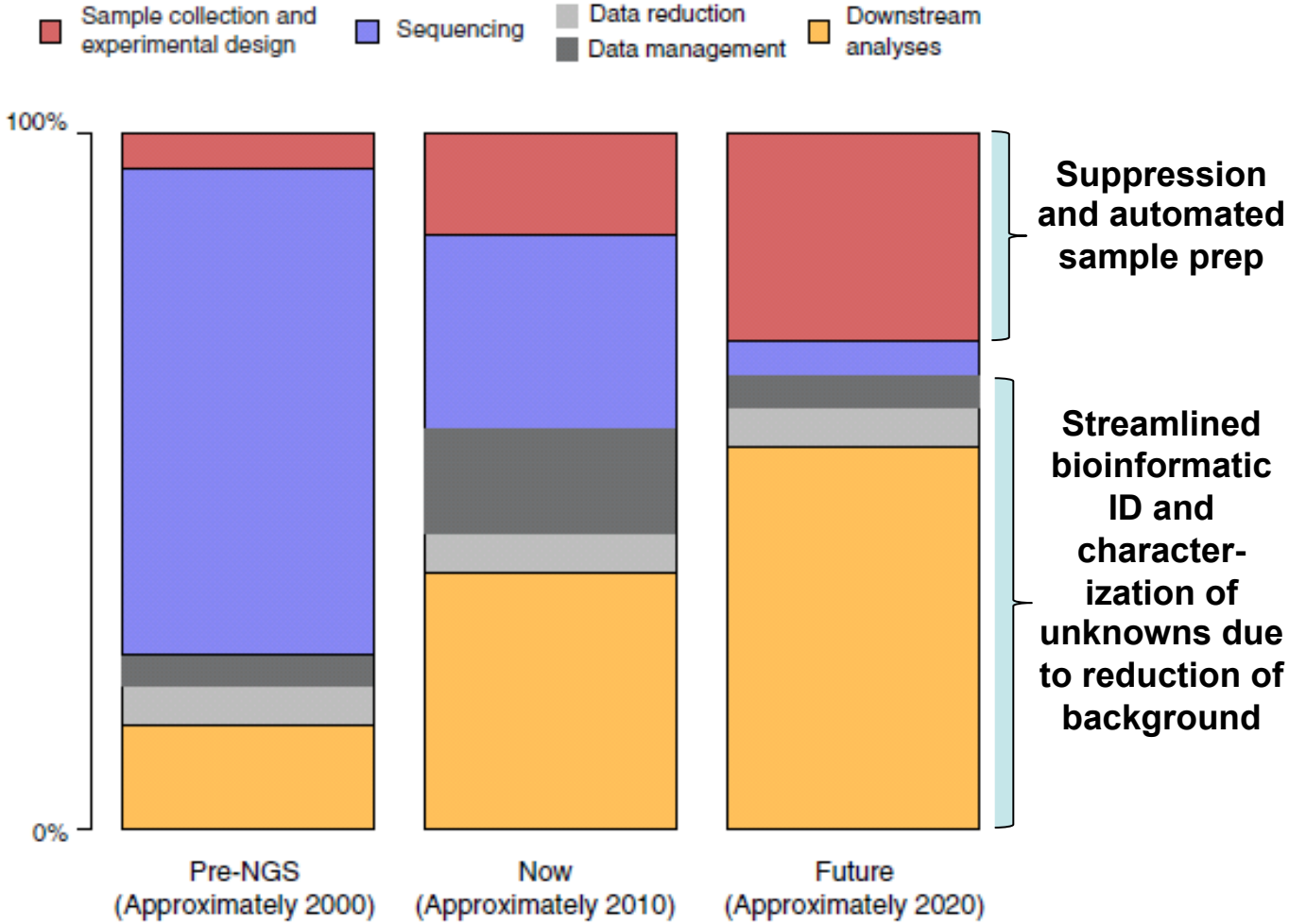
Deplete non-informative NA to improve efficiency of SGS analysis

Increasing pathogen signal to noise can be accomplished through suppression



The RapTOR approach was designed to address future challenges of SGS-based sample analysis





Sboner *et al. Genome Biology 2011, 12:125*

10 -100 X suppression would have significant effects on real samples

RNA viruses: The genomes are roughly 10kb, so
To get to 2x coverage, (some contigs): 200 reads at 100bp
to get 10x coverage (high confidence contigs):1000 reads at 100bp.
100x coverage (major chunks of genome) 10000 reads at 100bp.

For Bacteria with 2Mb genome

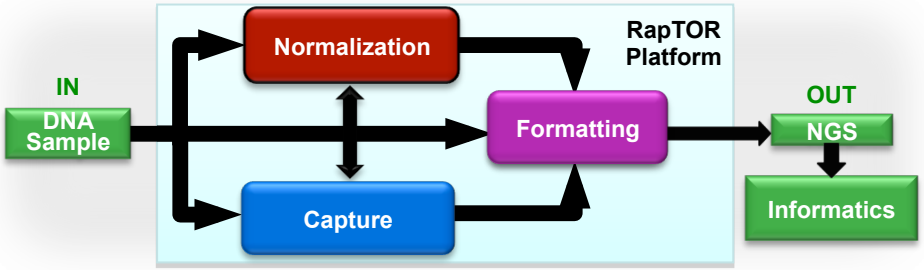
To get 10x coverage at 100bp is 200K reads.
To get 100x coverage of course would take 2M reads,

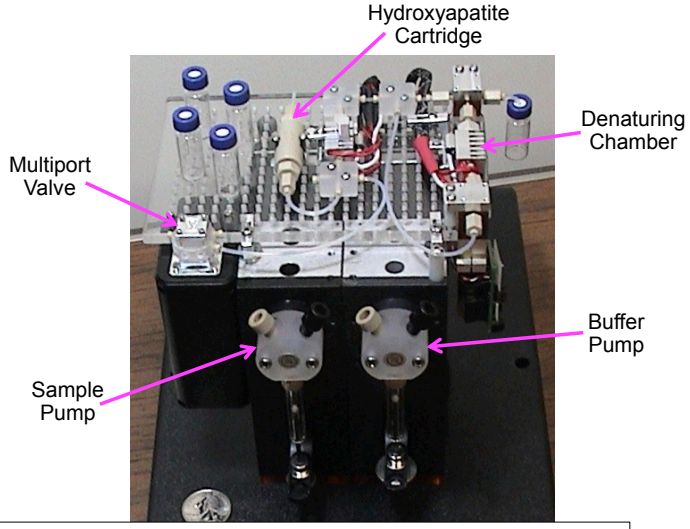
Viruses:
Many of the HCV samples have a few K hits, so 10x suppression would allow you to sequence the genomes well (modulo scaffolding issues).
Many of them also had hepatitis B at a few tens to a few K reads.
HBV as a small genome of ~3kb so 10x suppression yields 30X + coverage.

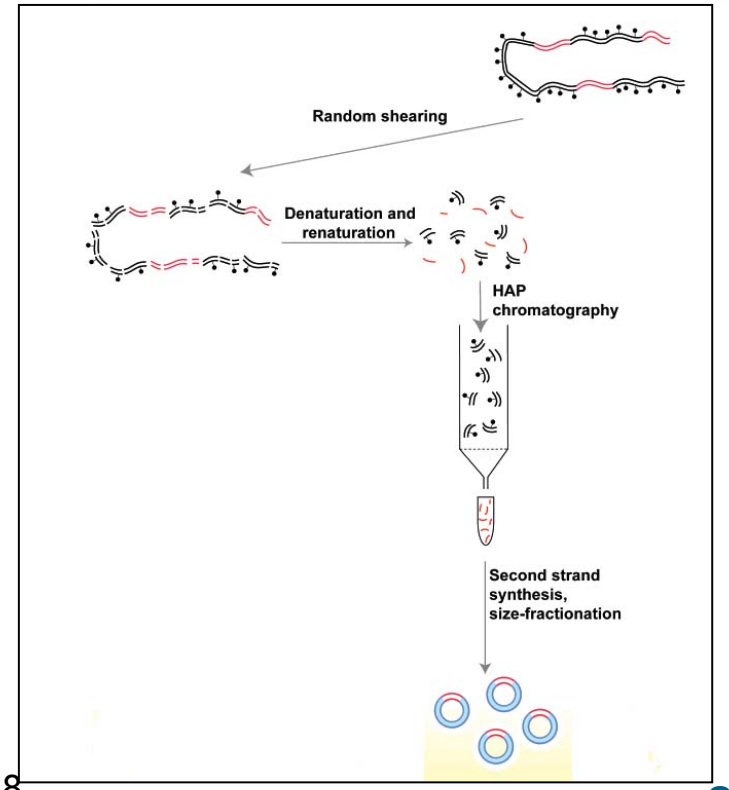
Bacteria:
Samples with high counts: tens of K to low 100K.
10x suppression yields 20x coverage

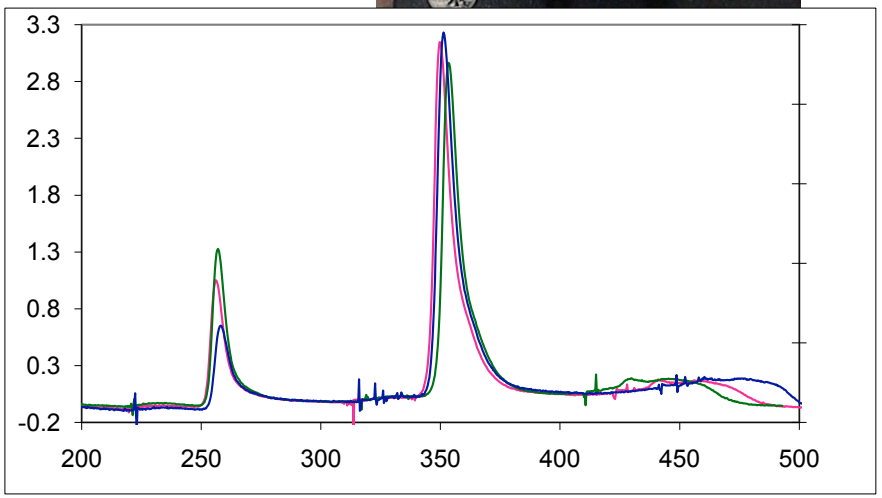
Samples with Low counts, between 1500 and 10K counts.
10x suppression yields 1-5x,coverage yields some contigs, species info and possibly virulence islands.
100x suppression would be desirable, but you could do a lot with 20-50x suppression.

HAC normalization has been implemented in benchtop and automated systems



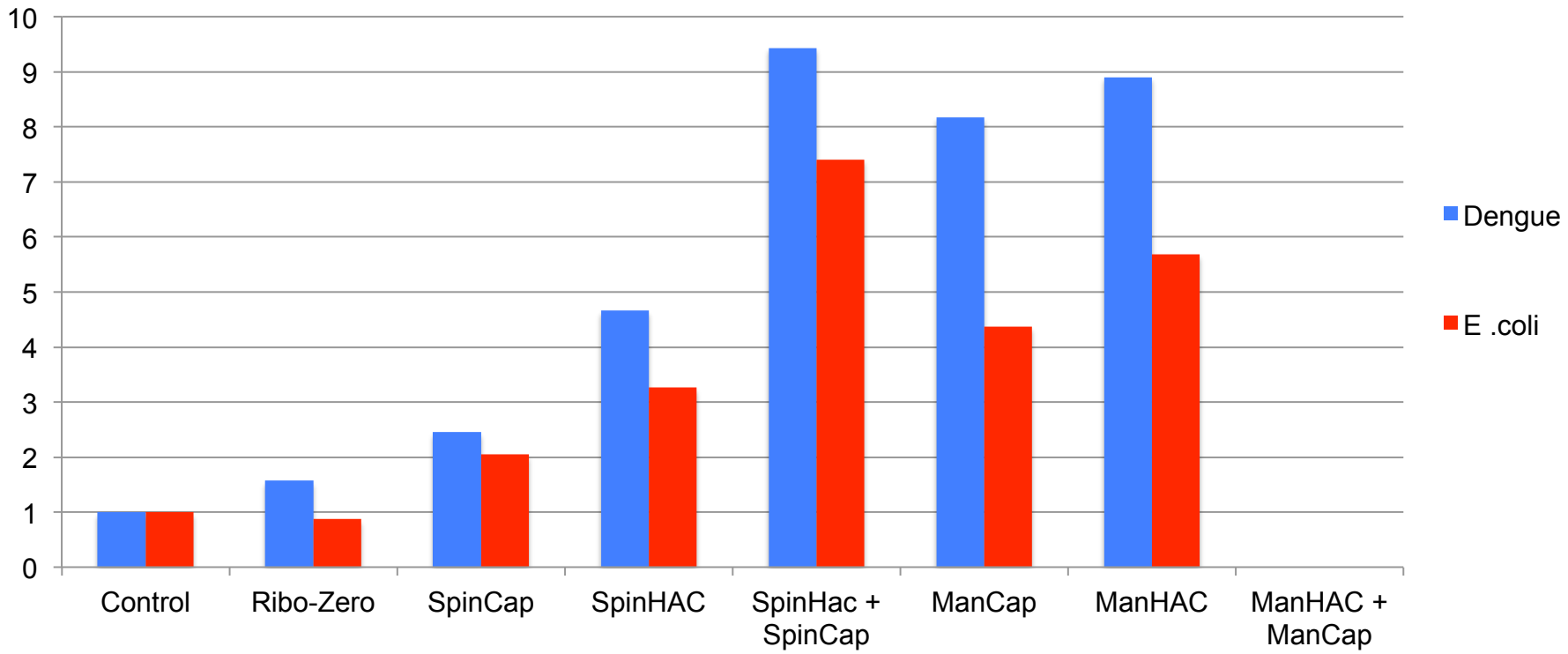






Sandia National Laboratories

Suppression increases hits on pathogens in mock clinical samples



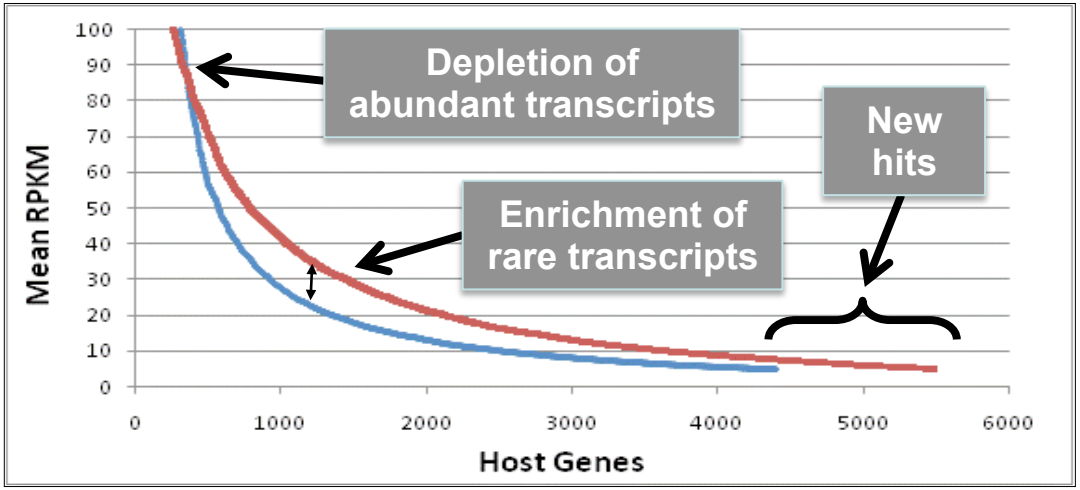
Multiple manifestations of the technique:

- Benchtop spin methods of normalization “HAC” and capture
- Manually operated capillary chromatography based methods
- Automated and multiplexed capillary based methods

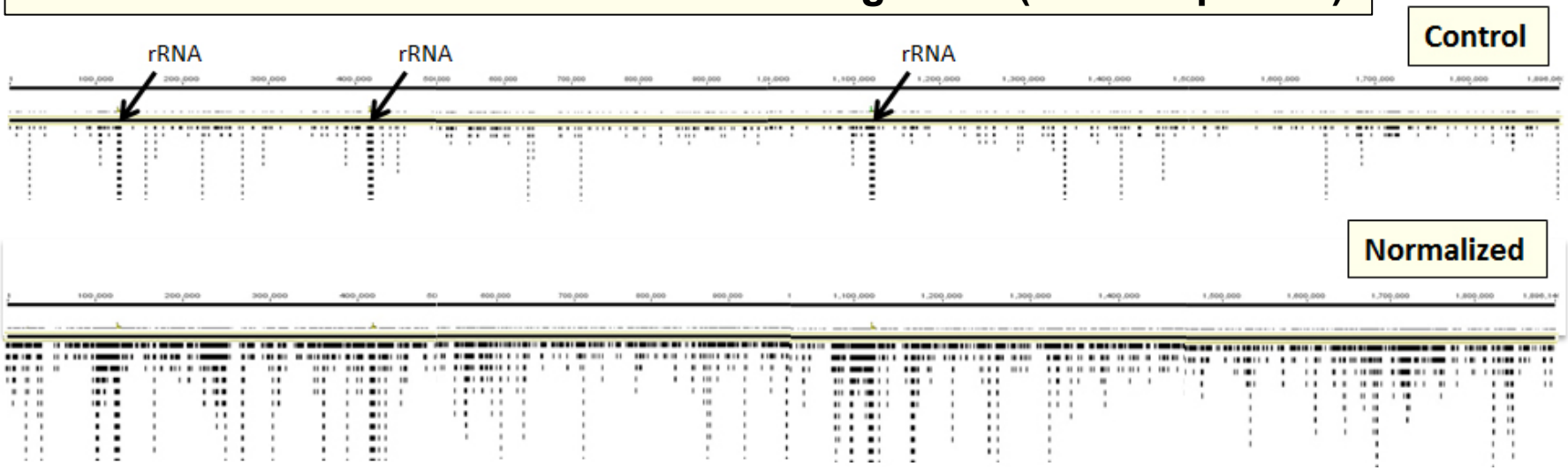
Suppression enriches pathogen sequences and increases coverage

Suppression enriches for rare nucleic acids, including those derived from the pathogen

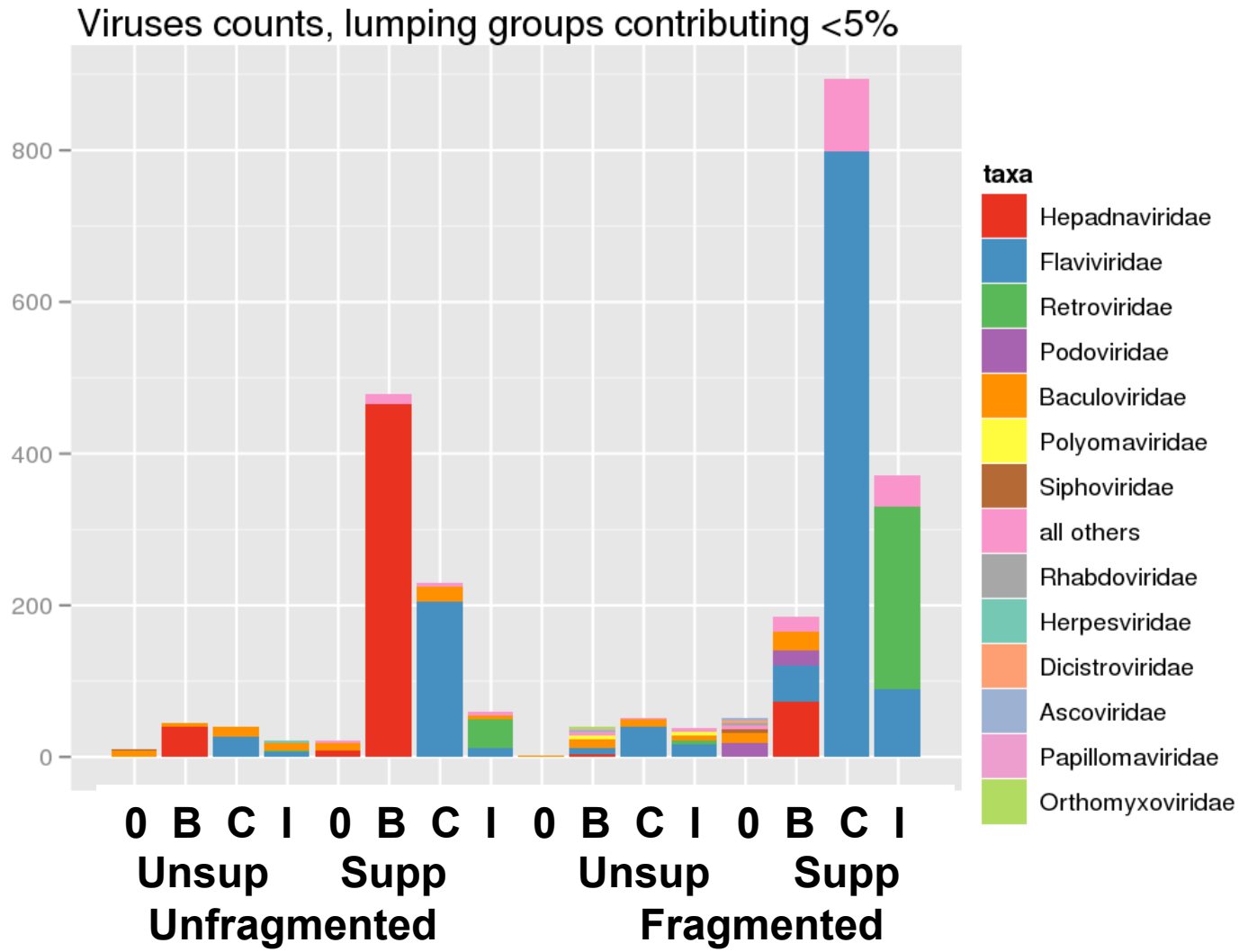
- Increases efficiency of SGS analysis
- Improves sequence coverage of pathogen



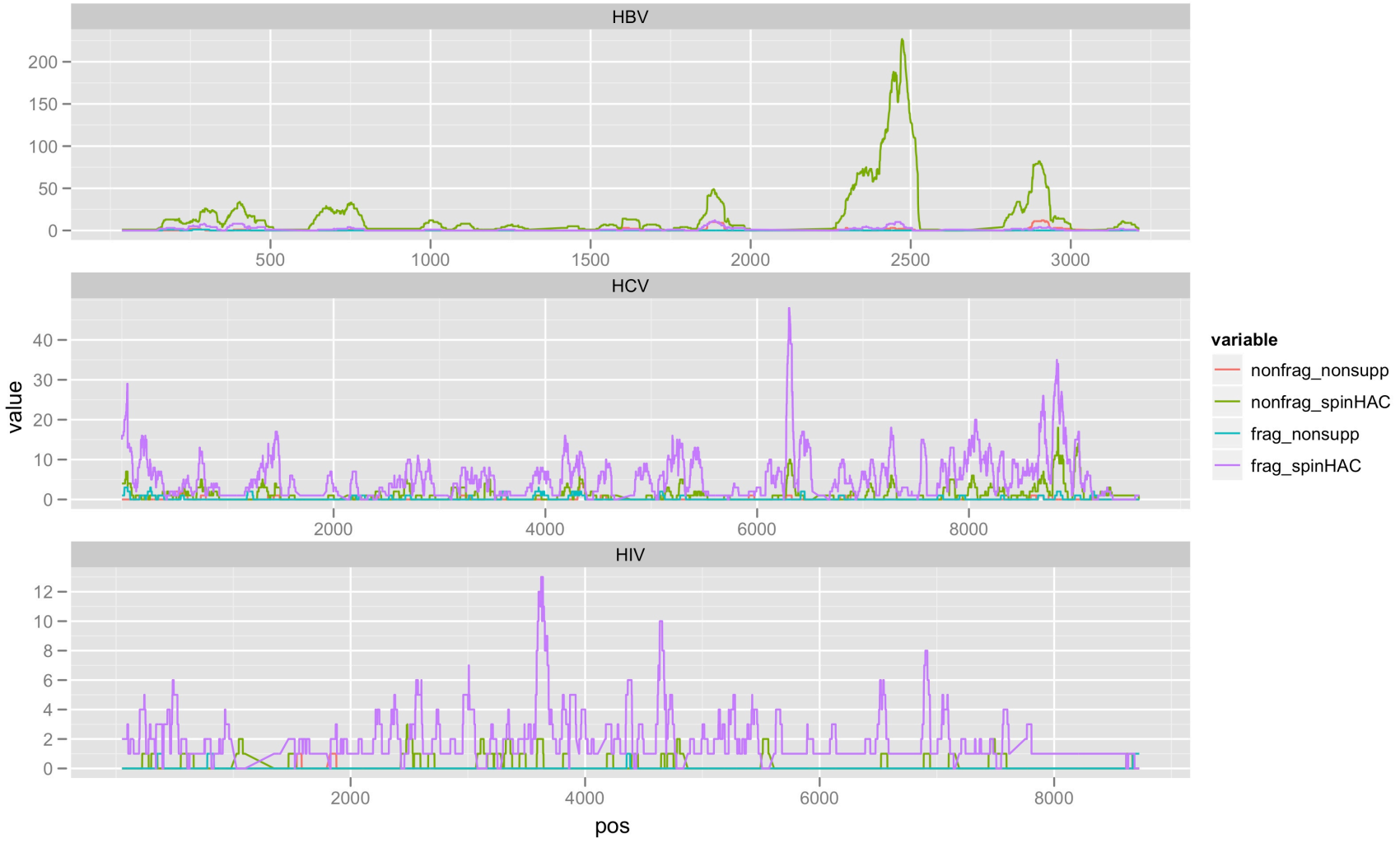
SGS "hits" on *F. tularensis* in human background (1:1000 spike-in)



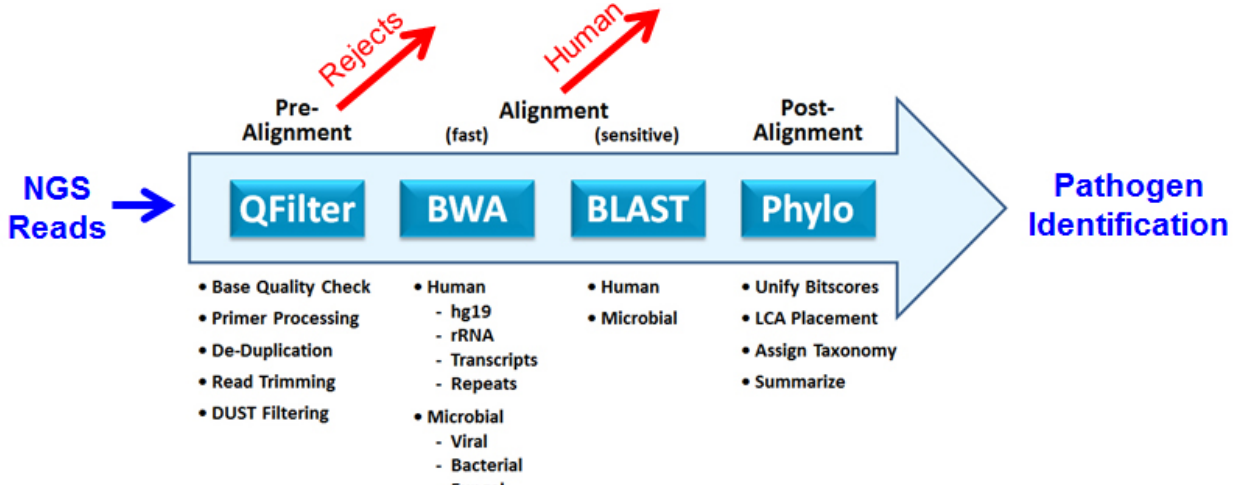
Suppression increases miSeq hits on viral Pathogens in clinical samples: plasma

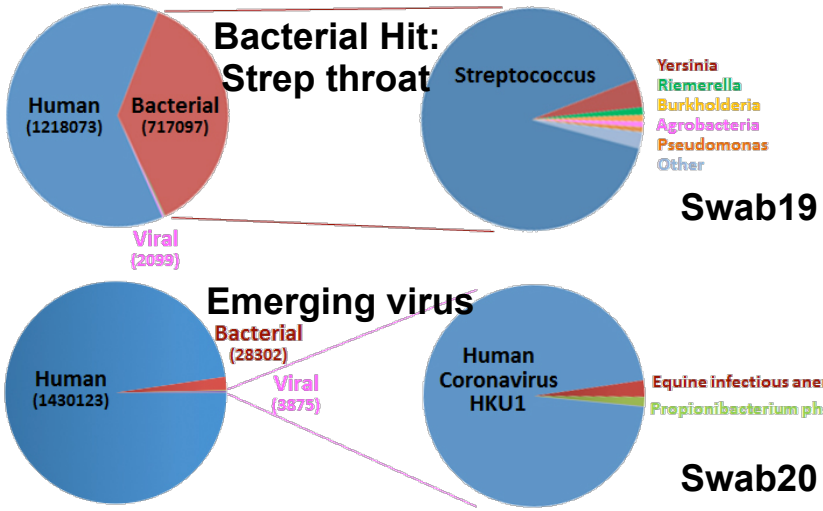


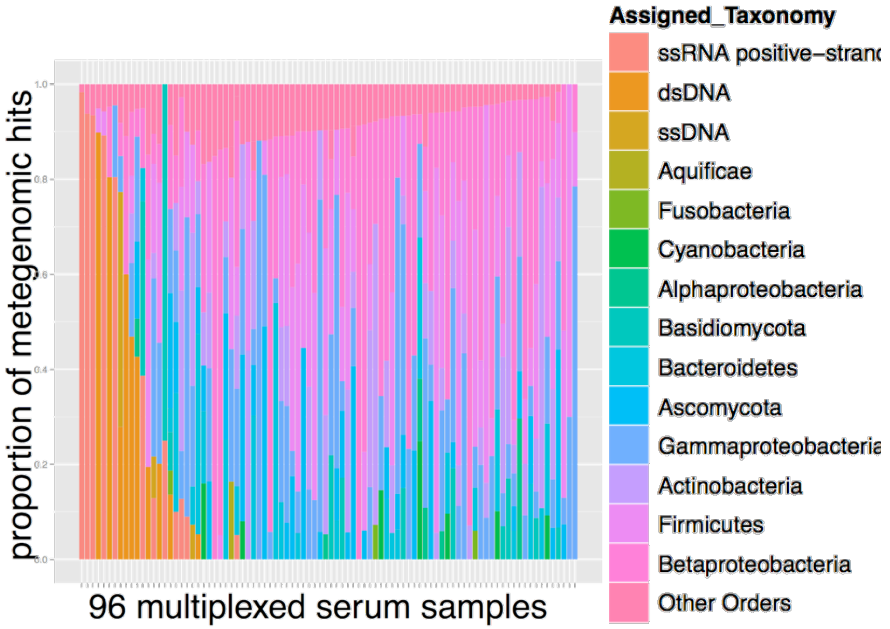
Increased coverage of viral genomes through suppression



Our bioinformatics pipeline enables efficient identification of pathogens in clinical samples







Nasopharygeal swabs from patients with respiratory complaints

SFAF June 7, 2012

Sandia National Laboratories

The difference between short term & sustained productivity is the problem

Short term areal production of 30-50 gm/m^2
–Commonly claimed

Long term areal production rates of 2 gm /m^2 (35 years of production)
–Ami Ben Amotz

Sub optimal growth conditions: Irradience, temp, salinity etc. and pond instabilities lead to lower long term production.

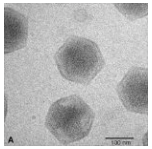
Closed PBRs can increase yield and postpone contamination, but have yet to be proved economically viable

Real time data on predator/pathogen load is enables proactive pond management

Presence of the biological agent can be necessary but not sufficient to crash

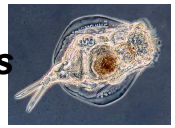
Agent + **Algae** + **Environment** = **Collapse**

Viruses

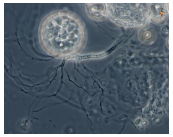


Bacteria

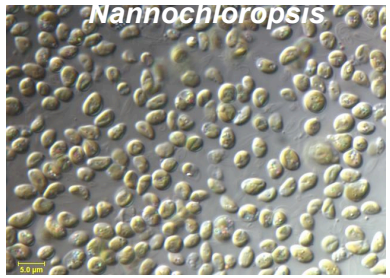
Predators



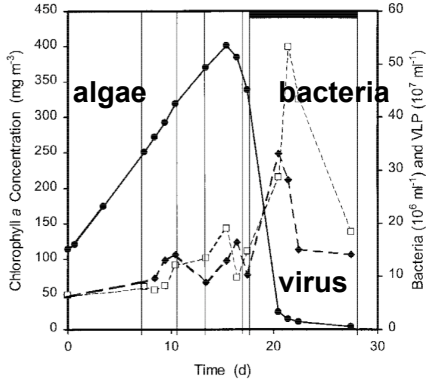
Fungi



Patterson & Laderman, 2001.



Environment (Temp, salinity, pH, CO2, nutrients)



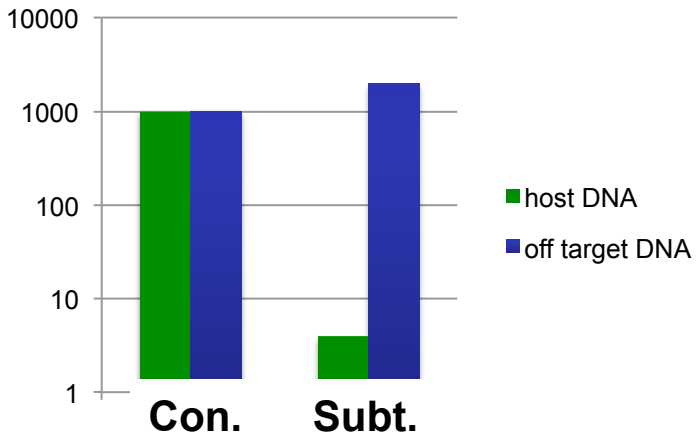
Herman Gons et al., Antonie van Leeuwenhoek, 81: 319-326, 2002.

"Perhaps the most worrisome component of the large-scale algal cultivation enterprise is the fact that algal predators and pathogens are both pervasive and little understood."

- DOE Draft Algal Biofuels Technology Roadmap (2009)

Parallels between this problem and that of the unknown biothreat were evid
We believed that elements of a RapTOR like approach would be effective.

Presumptive Identification of pond crash agents through sequencing



- ✓**Probe independent: detection of unknowns**
- ✓**Physical removal of host DNA prior to sequencing *up to 250 fold***
- ✓**No dependence on microscopic analysis/expertise**
- ✓**Effective for a wide variety of algal pathogens/predators**

% hits	Best hit	type
29	*Cercomonas plasmodialis;*	flagellate
21	*Aplanochytrium stocchinoi*	marine fungus
16	*Chaetonotus neptuni*	unsegmented worm-like
4	*Labyrinthuloides minuta;*	plant pathogen
4	*Platyreta germanica*	parasite/predator
4	*Amphora cf. capitellata*	competitor

Conclusions

We have developed molecular suppression techniques (HAC normalization, capture) that enable selective enrichment of pathogen NA in complex samples for targeted SGS analysis.

- **Suppression techniques implemented in benchtop (spin) and capillary chromatographic architecture**
- **in process of developing multiplexed automated capillary system**

We have demonstrated the effects of suppression on pathogen characterization in mock clinical and clinical samples

We have developed a metagenomics analysis pipeline that enables efficient identification of pathogen-derived NA sequences in complex clinical samples.

Acknowledgments

Steve Branda
Ken Patel
Joseph Schoeniger

Michael Bartsch
Zachary Bent
Bryan Carson
Joshua Crawford
Deanna Curtis
Hanyoup Kim
Sadie LaBauvre
Pamela Lane
Stan Langevin
Chi-Chi May
Robert Meagher
Milind Misra
Nicholas Pattengale
Martha Perez-Arriaga
Amy Powell
Ron Renzi
Bryce Ricken
Anupama Sinha
Owen Solberg
Numrin Thaitrong
Mary Tran-Gyamfi
Victoria VanderNoot
Kelly Williams